Shaking the Tree: An Agency-Theoretic Model of Asset Pricing*

Jamsheed Shorish[1] and Stephen E. Spear[2]

[1] Department of Economics and Finance, Institute for Advanced Studies, Vienna Austria shorish@ihs.ac.at
[2] Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh PA ss1f@andrew.cmu.edu

1 Introduction

The conditional heteroskedasticity and long-memory persistence apparent in financial times series is one of the most thoroughly studied but least understood of empirical regularities in financial economics. These regularities were first noted formally in the mid-1960s by researchers examining the long-run statistical behavior of stock prices, interest rates, and foreign exchange rates (see e.g. [8], [12], or [21]). In the early 1980's, the availability of high-speed computers made it possible for researchers to begin modeling the time-varying behavior of these time series explicitly. Early work on these topics includes the original Autoregressive Conditional Heteroskedasticity (ARCH) model of [7], and work by [17] and by [10] on fractional differencing (which built on the seminal analysis of fractional Brownian motions by [22]). Since that time, a number of extensions and elaborations of these models have appeared (see [2] for a detailed review of this literature). Application of these models to the stock price, interest rate, and foreign exchange data generally yields results which are highly statistically significant. It is no surprise, then, that ARCH and fractionally integrated time-series models have become increasingly more popular as tools for analyzing financial data.

What is surprising, however, is the relative dearth of theoretical results which might explain the observed time-varying volatility and/or persistence in the data. Standard asset pricing models of the type originally formulated by [20] or [3] are simply too stationary to deliver the desired effects. In the [20] framework, for example, the price of asset i is given by

$$p_i(y_t) = \beta E_t \left\{ M_i(y_t, y_{t+1}) \left(y_{t+1} + p_i(y_{t+1}) \right) \mid y_t \right\}$$

*We would like to thank Beth Shorish for her patience and guidance during this project, as well as conference participants at the 1998 North American Econometric Society Summer Meetings, Montreal, and the 53rd Econometric Society European Meetings, Berlin for their many useful comments.

where β is the discount rate, M_i is the pricing kernel, and y_t is the vector of dividends paid at time t. With a stationary pricing kernel, only the assumption of conditional heteroskedasticity in the dividend process will deliver ARCH effects in this framework. In models with production, convergence to steady-state equilibrium has the effect of eliminating non-stationarities in the model which might generate time-varying conditional volatilities, unless, of course, one assumes that the exogenously given shock process is itself ARCH. This approach has been examined explicitly in work by [1] and [11]. Each of these papers allows for underlying dividend processes with time-varying conditional variances and examines the implications of changing risk on the general equilibrium behavior of the model. While such studies can yield interesting insights, they hardly constitute a theoretical explanation of ARCH.

There have been several hypotheses put forth to explain ARCH effects. One such hypothesis holds that serially correlated news arrival drives increases in variance (see, for example, [5] or [9] for details). While this is plausible on its face, is poses the obvious question as to why the information arrival process should be serially correlated. Other researchers ([26], [27]) have examined time deformations that can occur when calendar time and market time proceed at different rates. In these models, trade can occur more or less frequently in the same calendar period. This can lead to observed volatilities which vary with calendar time even though the asset is covariance stationary when measured in "operational time". While these models do exhibit the desired conditional heteroskedasticity, they do not explain what might drive the fluctuations in trading rates that give rise to the time deformation to begin with.

A third hypothesis has been put forward by [14] to explain the observed ARCH effects in interest rate time series. In their model, market incompleteness together with constraints on borrowing leads to differences in savings behavior between rich and poor agents. This implies that the distribution of wealth in the economy influences the degree to which otherwise nicely behaved stochastic endowment shocks affect the interest rate in the model. Time variation in the distribution of income translates into time variation in the volatility of interest rates. While this mechanism may be at work in generating ARCH effects in interest rates, it seems implausible that variations in the wealth distribution could occur with sufficient frequency to drive the ARCH effects observed in stock prices.

The literature on theoretical explanations of the observed persistence in economics time-series is even sparser. The earliest result is work by [13] showing how aggregation of independent AR(1) processes can generate a process which is fractionally integrated. Since many financial and economic decision problems in stochastic environments lead to decision rules and pricing relations which are autoregressive, this would seem a plausible explanation for why economic aggregates exhibit persistence. One problem with this result, however, is that it is driven by the requirement that there exist AR(1) processes which have autoregressive parameters very close to 1. This requirement is generally not consistent with observed microeconomic data. A second problem is that the result isn't applicable to much of the financial data which is available in disaggregate form. More recent work on this issue has been done by [6] who examines a model based on the Ising models of physics. In Durlauf's model, firms are located on a lattice, and the activities of neighboring firms "spill over" and influence the original firm's production activities.

In this paper, we examine a very different mechanism based on an agency theoretic model of the internal dynamics of the firm. One of the weaknesses of exchange models of asset pricing is the need to exogenously specify the dividend process. This weakness is remedied somewhat in models of capital accumulation, since production is modeled explicitly. But the neoclassical view of the firm as a shell housing the basic technical processes that transform inputs into outputs fails to capture several essential aspects of real firms that may be relevant in determining the value of the firm. A key feature, which we focus on here, is the ability of a firm's managers to respond to favorable opportunities by expanding output (or sales from inventory) and to reduce output (or restock inventory) in response to unfavorable shocks.

We examine this feature using a standard principal-agent model together with a particular reduced form for production which allows the agent to control, at some cost, the rate of growth of the firm's output. To keep the analysis relatively simple, we embed this part of the model in a simple variant of the [20] asset-pricing model. As in Lucas, we have a single representative agent who owns an orchard full of trees (the assets) which bear stochastic amounts of completely perishable fruit (dividends) in every period. Unlike Lucas, however, we do not assume that the process of fruit production is completely exogenous. Instead, we include an agent, who one may think of as the gardener, who can influence the production of fruit by exerting effort in the orchard. Specifically, we assume that the agent can cause the amount of fruit produced to grow by exerting effort. In the context of the tree model, we can think of our gardener's fixed amount of time being allocated between routine crop tending and innovations in fertilizers, root stocks, and hybridization. The latter facilitate growth, which in turn demands more of the former activities, making it more difficult to engage in activities that promote innovation.

In order to keep the analysis relatively simple, we will focus on the case of a single asset or firm. The owner of the firm contracts with a manager to operate the firm and pays him a portion of the output as compensation. Again, for simplicity, we assume that the principal's interest is in maximizing the value of the firm, while the agent is risk averse with respect to his income. We consider a simple repeated relationship between the principal and agent without commitment. Under these assumptions, we show that the optimal contract generates a time-series for the firm's price which exhibits significant conditional heteroskedasticity and long-memory persistence, even when the underlying innovations are i.i.d.

In Section 2 we lay out the formal model. In Section 3 we look at a specific parametrization of the model and indicate how to solve this model numerically. Section 4 compares the first-best and second-best contract equilibria for the model, using both analytical conclusions and the numerical solution. Section 5 examines simulated time-series data generated by the model, and tests the data for evidence of long-memory persistence and of ARCH effects. We find that for certain specifications of the manager's preferences, both of these phenomena occur. Section 6 presents conclusions, while an Appendix contains the solution of the first-best contract, as well as the existence proof for the second-best contract equilibrium and details of the numerical simulation.

2 The Model

The model is based on the stochastic asset pricing model of [20], in which the firm is comprised of two agents, an owner and a manager. The manager's action is defined as an effort level which contributes to the production of the firm. Following Lucas, we associate the firm's production with a simple specification of a dividend process:

$$x_t \sim G(x_t \,|\, x_{t-1}, a_t) \tag{1}$$

where x_t and x_{t-1} are the current and previous dividends, respectively, a_t is the manager's action, and G is a continuous conditional distribution function over a compact support X—for simplicity we specify a first order dividend process.

Each period the manager chooses an action which maximizes the expected utility for the upcoming period. We assume that the manager is accepting a series of one-period contracts, and cannot commit to a longer contract. The manager's preferences are separable and given by

$$U(w_t, a_t) \equiv u(w_t) - h(a_t) \tag{2}$$

where w_t is the manager's wealth, defined as the wage paid by the owner, u is an increasing, continuous concave function, and h is a strictly increasing, continuous convex function. The h function measures the disutility of working to the manager. Finally, we assume that the agent possesses outside employment opportunities, which sets a lower bound $\bar{u}$ for the utility the agent must receive from employment.

The owner of the firm wishes to maximize the firm's value, i.e. the asset price. The owner must pay the manager a wage w_t in return for the manager's labor a_t in production (which is summarized by the dividend process). We suppose that the manager's labor contribution is unobserved when the wage is specified–thus the owner can only condition the wage upon the current dividend x_t. The owner seeks to

$$\max_{a_t, w_t} E_{t-1}(p_t), \tag{3}$$

where E_j denotes the conditional expectation operator with respect to information known at time j.

As in [20] the price of the firm is given by the discounted expected future dividends, net of the labor wage. This can be succinctly written as

$$p_t = \beta E_t p_{t+1} + x_t - w_t(x_t), \tag{4}$$

where $\beta \in (0, 1)$ is the owner's discount factor.

To complete the model, some mechanism for the owner's expectation of the future price must be defined. Suppose that the owner uses prior values of the state variables, in this case the dividend, in order to form her expectations. We assume that only the expected current dividend value is used in the forecast, and that the forecast function $v : X \to \mathbb{R}_+$ is continuous and time invariant:

$$E_t(p_{t+1}) = v(x_t). \tag{5}$$

The owner's problem is to select an action a_t and a wage w_t to maximize the value of the firm, given that the action of the agent will be unobservable. Using equations (1) to (5) we can restate the owner's problem as

$$\max_{a_t, w_t} \int_X (x - w_t(x))\, dG(x\,|x_{t-1},\, a_t) + \beta \int_X v(x)\, dG(x\,|x_{t-1},\, a_t) \tag{6}$$

such that

$$a_t \in \arg\max_a \int_X u(w_t) dG(x \mid x_{t-1}, a) - h(a), \tag{7}$$

$$\int_X u(w_t)\, dG(x\,|x_{t-1},\, a_t) \;-\; h(a_t) \geq \bar{u}.$$

Note that in this form the owner faces a typical second-best repeated static moral hazard problem–the dividend process is conditioned on the manager's choice of labor, but this is not incorporated into the contracting institution by the agents. While the single-period contract specification is used for reasons of tractability, it may be helpful to think of this as describing a firm which hires and fires labor of the same type every period. A newly-hired manager may observe that previous labor has affected the dividend process (as the distribution function for dividends is common knowledge) but has no control over the previous manager's labor choice.

We would like to examine the solution to the owner's problem without having to worry about the learning dynamics associated with the forecast function v. That is, we will assume that the owner has already learned the rational expectations equilibrium (REE) price function, and uses this function to forecast future prices. In other words, the owner knows the actual function v^* which takes the observed dividend and returns the expected value of the firm, i.e. $E_{t-1} p_t \equiv v^*(x_{t-1})$. This means that the REE price function must satisfy

$$\begin{aligned} v^*(x_{t-1}) = \max_{a_t, w_t} \Bigg\{ & E_{t-1}\left[x_t\,|x_{t-1}, a_t\right] - \int_X w_t(x)\, dG(x\,|x_{t-1},\, a_t) \\ & + \beta \int_X v^*(x)\, dG(x\,|x_{t-1},\, a_t) \Bigg\}. \end{aligned} \tag{8}$$

such that

$$a_t \in \arg\max_a \int_X u(w_t)\, dG(x\,|x_{t-1},\, a) \;-\; h(a), \tag{9}$$

$$\int_X u(w_t)\, dG(x\,|x_{t-1},\, a_t) \;-\; h(a_t) \geq \bar{u}. \tag{10}$$

In the Appendix it is shown that under the current assumptions on the dividend process and some additional restrictions on preferences, there exists a function v^* which defines at each point in time the value of the firm. However, in order to more fully characterize the qualities of the value function (and the associated optimal labor and wage functions, respectively) it is necessary to introduce functional form assumptions on the dividend process and on preferences. These assumptions also allow for numerical simulations to take place. In selecting these functional forms we have attempted to maintain a balance between generality and tractability–even with the specifications given below, the rational expectations price function is still general enough that an analytical solution cannot be found, and numerical analysis must be performed.

Definition 1. *The dividend process is an AR(1) process with innovation in the mean, i.e.*

$$G(x_t \,|\, a_t, x_{t-1}) \Leftrightarrow x_t = a_t + \rho x_{t-1} + \varepsilon_t,\ |\rho| < 1,\ \varepsilon_t \sim N(0,1)\, \forall t.$$

Definition 2. $U(w_t, a_t) \equiv -\exp(-w_t) - \exp(a_t)$.

With Definition 2 we may adopt the first-order-approach (see e.g. [25], [18]) to replace the argmax operator in equation (7) with the associated first-order condition:

$$\frac{\partial}{\partial a} \int_X u(w_t(x))\, dG(x \,|\, a, x_{t-1}) - h(a) \bigg|_{a=a_t} = 0. \tag{11}$$

3 Numerical Approximation

The specifications for the dividend process and preferences are not enough to generate an analytical solution for the value, wage or labor functions. The approach taken in this paper is to leave the remaining functional forms as general as possible, and to instead focus on numerical solutions to the REE condition. The aim here is to identify and analyze the dynamics of the price and dividend processes given the numerical solution, instead of concentrating solely upon the analytical results of e.g. a local quadratic approximation of the value function v^*.

However, we would like to be able to compare the resulting owner-manager contract and, ultimately, the dynamics of the price process with a benchmark case. We can then see what this type of environment is ‘bringing to the table’ when compared with other contracting forms. In particular, it is interesting to compare the ‘second-best’ model outlined above with the ‘first-best’ problem, in which the manager’s action is set by the owner without considering the manager’s optimal choice. In this case the incentive compatibility condition is absent from the owner’s optimization, and the only thing the owner need worry about is giving the manager enough utility (in this case $\bar{u}$) to choose employment.

In the first-best case, the problem facing the owner is

$$\max_{a_t, w_t} \left\{ E_{t-1}\left[x_t \,|x_{t-1}, a_t\right] - \int_X w_t(x)\, dG(x\,|x_{t-1},\, a_t) \right.$$
$$\left. + \beta \int_X v(x)\, dG(x\,|x_{t-1},\, a_t)) \right\}$$

such that

$$\int_X u(w_t)\, dG(x\,|x_{t-1},\, a_t) \;-\; h(a_t) \geq \bar{u}.$$

In the Appendix it is shown that when Definitions 1 and 2 hold the wage function has the usual property that the manager’s risk is entirely smoothed away–regardless of the observed dividend, he always receives a utility of $\bar{u}$. Both the wage function and the effort function are constant, and the expected price function v^* is linear and increasing. Further details on the first-best contract will be presented when compared with the second-best contract below.

Unfortunately, in the second-best case it is not possible to find a closed-form solution. So there remains the problem of finding the expected price function (or ‘value function’) v^* of the firm. We adopt here a numerical approximation technique to identify the value and policy functions. The method of obtaining the value function used here is by iterating on a functional operator T, defined by

$$T\left(v^n\right)(x_{t-1}) = E_{t-1}\left[x_t \,|x_{t-1}, a_t^n\right] - \int_X w_t^n(x)\, dG(x\,|x_{t-1},\, a_t^n) \qquad (12)$$
$$+ \beta \int_X v^n(x)\, dG(x\,|x_{t-1},\, a_t^n).$$

where w_t^n and a_t^n are the optimal choices given the candidate value function v^n. Since T is a contraction (see Appendix) it follows that

$$\lim_{n\to\infty} T^n(v_0) = v^* \,\forall v_0 \in C_X^1. \qquad (13)$$

This condition simply states that for any initial continuous (bounded) function taking dividends into prices, iterations of the operator T on the initial function will converge to the rational expectations price function.

In the approximations these iterations are derived from a class of functions known as universal approximators. A universal approximator has the property that given a finite collection of points from the domain and range of an unknown function, the approximator can update its parameters such that it converges almost everywhere to the unknown function. Members of the class of universal approximators include neural networks, which specify a 'general' functional form up to a finite set of parameters. These parameters are then updated by iteration using the collection of points from the unknown function until they approach a set of 'true' parameters which in principle allow the neural network to arbitrarily approximate the unknown function. (See [16] for a discussion of universal approximation and neural networks, and [29], [19] for a general discussion of the applicability of neural networks to both functional approximation and regression analysis.) The properties of neural networks are by now well established–for the purpose of this paper, they are essentially a convenient method of implementing nonlinear least squares regression in a deterministic setting.

We are now in a position to outline the algorithm for numerically computing the rational expectations value function v^*:

1. Select an initial value function v^0. Using neural networks, this amounts to defining a network whose parameters are randomized.
2. Specify a finite grid over the dividend space X. Given v^0 and a distribution for the error process ε_t, numerically compute for each point in the grid the optimal values for w_t^0 and a_t^0 (i.e. carry out the optimization on the right-hand side of equation 12, where $n = 0$). For each grid point, the value for $T \circ v^0$ is computed.
3. Iterate the value function to $v^1 = T \circ v^0$, for which there is a finite collection of values given by step 2. These values define a neural network f^1 which approximates the function v^1.
4. Use the neural network approximation f^1 in place of v^0 in step 2. Repeat steps 2-3 until $\left\| f^n - f^{n-1} \right\| < \eta$, where η is some predefined error tolerance level. Call f^n the rational expectations value function, or v^*.
5. Given each point in the dividend grid, perform the optimization on the RHS of equation (12) using v^*. This gives in the optimal values for the wage and effort level for each grid point. These values serve to define a neural network $[w_t^*, a_t^*] = [w^*(x_t, x_{t-1}), a^*(x_{t-1})]$, which is the 'policy function' for the economy. Given the state of the economy (i.e., a realization of the current dividend, plus the previous dividend) the policy function tells the principal what wage should be paid, and how hard the manager should work in the current period.

The rational expectations value function is the law of motion for the economy. Once found, v^* defines the optimal a_t^*–this, combined with the previous level of the dividend x_{t-1} and a new realization of the error process ε_t, generates the current dividend x_t. Having observed x_t, v^* also defines w_t^*–this is the wage paid to the agent by the owner and depends only upon $[x_t, x_{t-1}]$ since the owner cannot observe the manager's effort. The actual price of the asset p_t is then

$$p_t = \beta v^*(x_t) + x_t - w^*(x_t, x_{t-1}). \tag{14}$$

With the current dividend x_t the next effort level a^*_{t+1} can be defined, and the process continues. Thus, the economy can be simulated and sequences of dividends and prices can be generated and analyzed.

Particulars on the specification of the networks $v^*(x_{t-1})$ and $[w^*(x_t, x_{t-1}), a^*(x_{t-1})]$, including the number of hidden units, number of iterations, convergence criteria, and other parameter values may be found in the Appendix.

4 First-Best and Second-Best Comparison

We start with the analytical 'benchmark' results for the first-best case. In the Appendix it is shown that the optimal value and policy functions are:

$$E_{t-1}p_t = v^*(x_{t-1}) = \frac{1}{1-\beta}\left(\frac{a^*}{1-\beta\rho} + \ln(|\overline{u}| - e^{a^*})\right) + \frac{\rho}{1-\beta\rho}x_{t-1}, \quad (15)$$

$$a^* = \ln\left(\frac{|\overline{u}|}{2-\beta\rho}\right),$$

$$w^* = -\ln\left(|\overline{u}|\left(\frac{1-\beta\rho}{2-\beta\rho}\right)\right).$$

Figure 1 displays the optimal value function for the case where $\bar{u} = -2$, $\beta = 0.9$, and $\rho = 0.9$. These are the identical parameter values used for the second-best approximation.

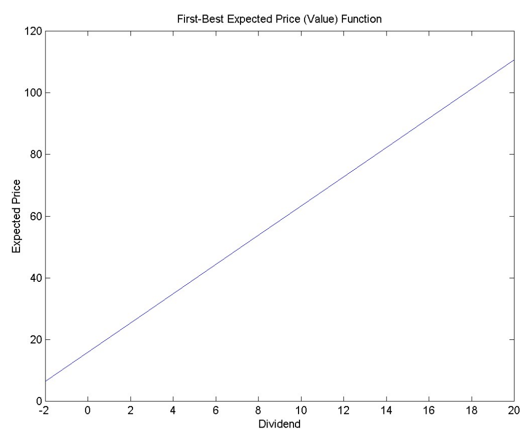

Fig. 1. First-Best Contract Expected Price

Clearly, in the first best case we will have no conditional heteroskedasticity or long memory–the actual price follows the process

$$p_t = x_t + \ln\left(|\overline{u}|\left(\frac{1-\beta\rho}{2-\beta\rho}\right)\right) + \beta\frac{1}{1-\beta}\left(\frac{a^*}{1-\beta\rho} + \ln(|\overline{u}| - e^{a^*})\right) + \frac{\rho}{1-\beta\rho}x_t \Leftrightarrow$$

$$p_t = c + \frac{1-\beta\rho+\rho}{1-\beta\rho}x_t, \tag{16}$$

where we have grouped constant terms into c for convenience. Since

$$x_t = a^* + \rho x_{t-1} + \varepsilon_t$$

we can lag and substitute (16) into the dividend relation to yield

$$p_t = \rho p_{t-1} + (1-\rho)c + \left(\frac{1-\beta\rho+\rho}{1-\beta\rho}\right)(a^* + \varepsilon_t).$$

The price thus follows an AR(1) process with the same autoregressive parameter as the dividend process, but with a different mean and variance. This result supports the intuition that because in the first-best case there is no response by the agent to changes in production, there should be no resultant correlation between production (or price) volatility from period to period. In the first-best case, the sole connection between the present and the past is the autoregressive dividend process.

In the second-best case, however, things are markedly different. Figures 2-4 present the numerical results for the second-best value function, wage function and effort function respectively. Figure 2 shows the numerical approximation of the value function given by the program (8)-(10), using Definitions 1 and 2 and the parameter values $\bar{u} = -2$, $\beta = 0.9$, and $\rho = 0.9$. The value function is strictly concave, and for high levels of the dividend the expected future price is nearly constant. This reflects the fact that for very high levels of the dividend the manager shirks a great deal, absorbing any expected future gains from the dividend. Figure 3 presents the manager's effort function–note that for low or negative dividend levels the manager wishes to work a (small) positive amount, but that as dividends rise the manager rapidly works less and less. Figure 4 presents the optimal wage as a function of the current and past dividends. From this we see that the manager receives positive compensation when there is a large positive gain in the dividend process (indicating hard work by the agent). Compensation then falls as the difference between the two dividends falls, and becomes sharply negative when the current dividend is far below the previous one.

5 Simulation of Time Series

Once the value function and optimal wage and effort functions have been approximated it is possible to simulate the economy and generate synthetic time series for analysis. The simulation of time series data is important because the second-best model has no analytical solution. Thus, analytical tools such

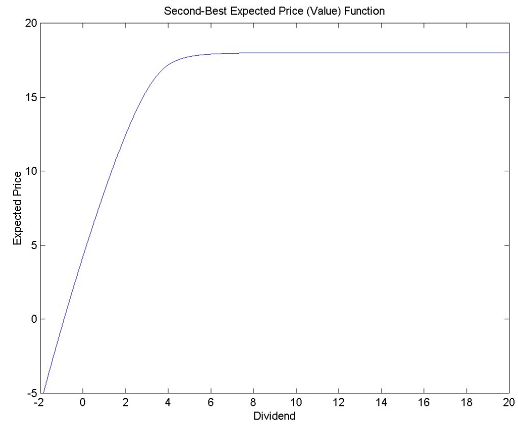

Fig. 2. Second-Best Contract Expected Price

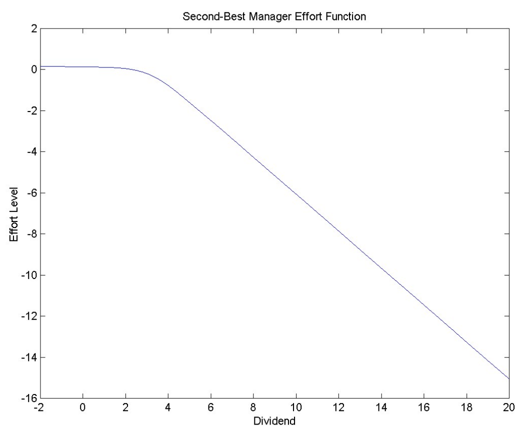

Fig. 3. Second-Best Contract Manager Effort

as comparative statics must be traded for tools which take advantage of the large number of simulated data points that can be generated from the model. Tools such as regression analysis do, of course, make the tacit assumption that the model is an accurate representation of the salient features of the owner-manager relationship in a real economy. Nonetheless, the complexity of the model and the relative paucity of relevant data in the real world are both strong incentives to use simulated time series data as a proxy for economic data generated from actual owner-manager behavior.

The owner-manager model is by its nature a nonlinear model of pricing. Thus, one might expect the dynamics of observed price and dividend sequences to reflect this nonlinearity. Of particular interest is the extent to

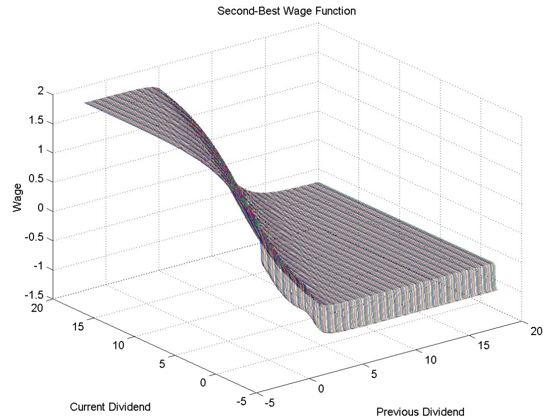

Fig. 4. Second-Best Contract Wage

which the nonlinearity in the model is capable of generating time-series data with properties observed in real stock price data. Thus, in analyzing the simulated time-series, we look for evidence of conditional heteroskedasticity and long-memory persistence.

The tools used for the analysis of the simulated time series are the generalized ARCH (or GARCH) model of conditional heteroskedasticity (see [7] and [2]), the estimated power spectrum of the simulated time-series, and a test for long-memory persistence. The results of this analysis indicate that 1) the price series exhibits strong ARCH-like behavior, with the GARCH (1,1) model demonstrating significant correlated volatility, and 2) the power spectrum and the long-memory test indicate some long-memory persistence.

The second-best value function and the optimal wage and action functions were used to generate time series of both prices and dividends for one hundred thousand periods. A typical example of the dividend time series, along with the associated wage compensation and effort levels for 200 periods are given in Figures 5 and 6. A sample path of the second-best price for the same 200 periods is presented in Figure 7. Since it is the value of the firm which empirically demonstrates ARCH-like behavior and long-memory persistence, only the time series for the price of the firm was tested in what follows.

5.1 Spectral Analysis

To answer the question of whether the simulated price series exhibits persistence, we first examine the empirical power spectrum. We calculated the Welch-averaged spectral density of the simulated price series with 100,000 data points, with a Hanning window 100 units wide. The smoothed spectrum with 95% confidence bands is shown in Figure 8.

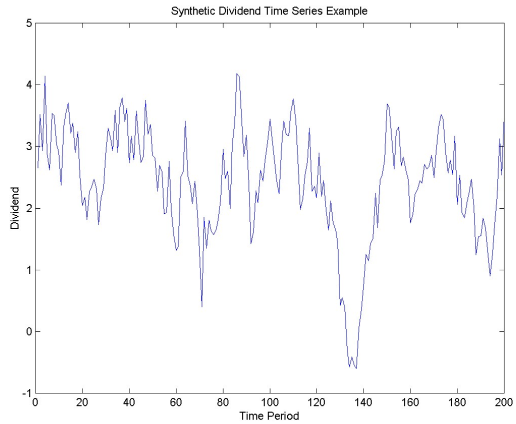

Fig. 5. Sample Dividend Series from Synthetic Data, 200 Periods

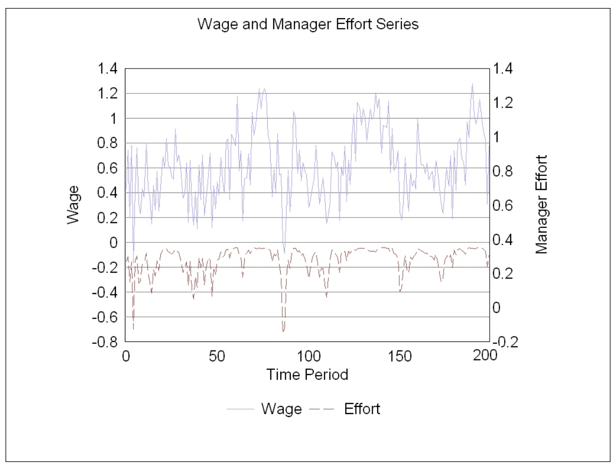

Fig. 6. Wage and Manager Effort Series from Synthetic Data

The spectrum indicates considerable power in the lower frequencies, a hallmark of long-memory processes. However, it does not appear that the spectrum demonstrates strong long-memory persistence. Rather, it appears at first blush that the spectrum of the second-best price series is still dominated by the AR(1) structure of the dividend process. In order to ascertain if long-memory has any impact we turn to a 'rough and ready' test of persistence given by [10]. This test attempts to verify whether or not the spectrum bears a some similarity to an inverse power law in frequency. In the literature on fractionally integrated time-series, this is often referred to as "1/f " noise.

To test the hypothesis that the observed data behave as 1/f noise, we regressed the log of the power spectrum against the log of the (sin of) fre-

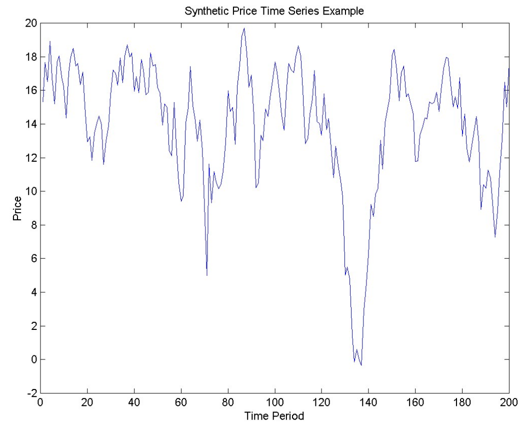

Fig. 7. Sample Price Series from Synthetic Data, 200 Periods

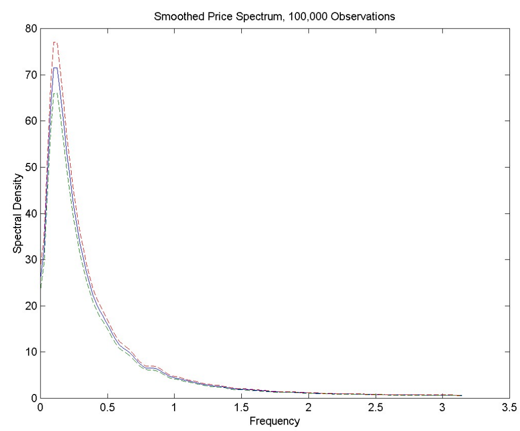

Fig. 8. Smoothed Price Spectrum with 95% Confidence Bands, 100,000 Obs.

quency, and the results are presented in Table 1. As shown in the table, the estimated long-memory exponent $d = 0.61$ is significant beyond the 99% confidence level. We conclude, then, that the simulated price series exhibits some long-memory persistence.

This results raises the obvious question as to what causes the observed persistence. One way of approaching this is to consider an experiment first performed in the 1940's by the hydrologist Harold E. Hurst. Hurst's experiment involved generating random sequences of biased random walks, and analyzing the properties of the resulting time-series. Subsequences were characterized by a fixed bias in the step size of the random walk. A second random variable determined when the bias would be changed, generating a new sub-

Table 1. Spectral Regression (Geweke & Porter-Hudak) Estimate Results

Variable	Coefficient	Std. Error	t-statistic
intercept	2.2096	0.0705	31.349^{**}
d	0.6088	0.0441	13.816^{**}
R^2	0.7578		

sequence. When Hurst examined large data sets generated in this fashion, he found significant evidence of long-memory persistence. Subsequent work on time-series models with exogenous structural breaks has confirmed the long-memory properties of the time-series generated by such models (see e.g. [24]). In our model, structural breaks in the trend of the stock price time series occur when the drift parameter swings from negative to positive (or vice-versa) in response to increased (or decreased) effort by the agent. These breaks occur endogenously (but randomly, given their dependence on the innovation process of the dividend) and, we believe, generate the observed persistence. As we will see in the following section, the trend breaks may also generate the significant conditional heteroskedasticity observed in the simulated price series.

5.2 Estimation of the GARCH (1,1) Model

The GARCH (1,1) model for conditional heteroskedasticity is defined as

$$p_t \mid \Psi_{t-1} \sim N\left(\beta \mathbf{p}_{t-l}, \sigma_t^2\right),$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \sigma_{t-1}^2,$$

$$\varepsilon_t = p_t - \beta \mathbf{p}_{t-l},$$

where Ψ_{t-1} is the information set available at time $t-1$, $\mathbf{p}_{t-l} \equiv [1, p_{t-1}, \dots p_{t-l}]'$ is an $(l+1)-$dimensional vector of lagged endogenous variables, ε_t is the residual of the mean regression, and β, $\alpha \equiv [\alpha_0, \alpha_1, \alpha_2]$ are $(1 \times l+1)$ and $(1 \times 3)-$dimensional vectors, respectively. This specification allows for the possibility of 'booms' and 'busts' (or alternatively, 'fads' and 'bubbles') in the price sequence, and has (along with its variants E-GARCH and N-ARCH) been implemented extensively on economic data (see e.g. [4], [15] and [23]).[3]

The GARCH (1,1) model of the second-best price series was estimated using Maximum Likelihood with 100,000 observations, with a first-order autoregressive process for the conditional price expectation. Before estimation,

[3] Veronesi (1996) also used the GARCH (1,1) specification to test for ARCH, in a model which uses 'regime shifts' (similar to the trend breaks of Perron [1989]) to generate correlated volatility in asset returns.

the Lagrange Multiplier (LM) test for ARCH ([7]) and the Jarque-Bera normality test were applied to the residuals of the mean equation. The results of the estimation and tests are presented in Table 2. The LM test strongly rejected the i.i.d. residual hypothesis at the 99% confidence level, while the Jarque-Bera test rejected the normality hypothesis beyond the 99% confidence level. These are strong indications that the time series contains some measure of correlated volatility.

Table 2. GARCH(1,1) Estimation Results

Variable	Coefficient	Std. Error	z-statistic
Mean Equation: $E_{t-1}p_t = \beta_0 + \beta_1 p_{t-1}$			
Constant	2.5293	0.0266	95.015**
p_{t-1}	0.8309	0.0018	453.23**
LM ARCH Test: $T * R^2 = 488.27$, Pr$(i.i.d.)$ $< 10e^{-6}$			
Jarque-Bera Statistic: 3815.3, Pr$(normal)$ $< 10e^{-6}$			
Variance Equation $E_{t-1}\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\sigma_{t-1}^2$			
Constant	0.4802	0.0243	19.739**
ε_{t-1}^2	0.0631	0.0025	25.561**
σ_{t-1}^2	0.7738	0.0097	79.328**

** refers to significance at the 99% confidence level.

The coefficients of the GARCH (1,1) model were all statistically significant beyond the 99% confidence level. In addition, the conditional variance process is strongly persistent (with $\alpha_2 = 0.774$). This provides reasonable grounds for acceptance of the GARCH (1,1) model as demonstrating the existence of conditional heteroskedasticity in price data which is generated by second-best contracting. Naturally, fitting the GARCH model to the data is an *a priori* model misspecification, since the price data are actually generated by (14). The GARCH estimation is in this case simply used to demonstrate the strong presence of correlated volatility in the price data.

6 Conclusion

We have seen that the simple model presented here yields a price sequence which can have hidden nonlinear behavior. In addition, it appears that examples of this sequence are fit well by an ARCH-like specification, which has been noted to exist in empirical data. These results support the conclusion that endogenous correlated volatility and persistence is possible when the manager influences the production process and is in turn affected by his contractual relationship with the owner.

The model of owner-manager behavior in an asset-pricing model presented here has been simplified in many key ways. First and most important, the type of contract that the manager is able to make with the owner is essentially static–commitment only occurs over the single period the contract is in force, and the same contract is assumed to be accepted in perpetuity. Further investigation would see whether a more general model, incorporating a multiperiod contract with commitment, would be substantively different from the simplified model presented here. It would be interesting to see, for example, whether the correlation between dividends and prices is less strong under a fully dynamic model, since it is empirically observed that dividend time series is less volatile than (hence, less strongly correlated with) the asset price series.

In addition, it is not clear whether the manager affects the mean of the production process (as assumed here), the variance of the process (as would be true, for instance, if the manager could influence the impact of the exogenous shock upon the dividend), or both. Future research will develop a production-based model of manager effort which can then suggest a dividend process such as equation (1) as a direct conclusion. Of course, given the analytical complexity of the simplified case, it is not at all clear that a more general model with the above extensions would yield analytically testable conclusions. Rather, it would appear from this presentation that the tools of numerical approximation and simulation of time series would be just as valuable for these cases.

7 Appendix

7.1 Solution of the First-Best Contract

Consider the first-best economy:

$$v^*(x_{t-1}) = \max_{w,a} \left\{ a + \rho x_{t-1} - \int w(x) dG(x|x_{t-1}, a) + \beta \int v^*(x) dG(x|x_{t-1}, a) \right\}$$

$$s.t. \int e^{-w(x)} dG(x|x_{t-1}, a) + e^a \leq |\overline{u}| \, ,$$

where we have used the functional forms from Definitions 1 and 2. From the optimization with respect to w we know that the wage payment will be independent of the observed output x, so that

$$e^{-w^*} + e^a = |\overline{u}| \Rightarrow$$

$$w^* = -\ln(|\overline{u}| - e^a),$$

given a particular level of managerial effort a.

The owner's optimization is now

$$v^*(x_{t-1}) = \max_a \left\{ a + \rho x_{t-1} + \ln(|\overline{u}| - e^a) + \beta \int v(x) dG(x|x_{t-1}, a) \right\}.$$

The objective function of the owner is concave in a. Posit a candidate policy function $a = a^*$, where a^* is some constant, and a candidate value function $v^*(x_{t-1}) = c_1 + c_2 x_{t-1}$. Then Bellman's equation at the optimal level of effort is

$$c_1 + c_2 x_{t-1} = a^* + \rho x_{t-1} + \ln(|\overline{u}| - e^{a^*}) + \beta c_1 + \beta c_2 (a^* + \rho x_{t-1}).$$

Matching coefficients yields

$$c_1 = a^* + \ln(|\overline{u}| - e^{a^*}) + \beta c_1 + \beta c_2 a^*$$

$$c_2 = \rho + \beta \rho c_2$$

or

$$c_1^* = \frac{1}{1-\beta} \left(\frac{a^*}{1-\beta\rho} + \ln(|\overline{u}| - e^{a^*}) \right),$$

$$c_2^* = \frac{\rho}{1-\beta\rho}$$

(note that this can also be directly verified by appealing to the equivalent sequence problem–cf. [28]).

The owner thus wishes to solve

$$\max_{a^* \in (-\infty, \ln(|\overline{u}|))} \left\{ a^* + \rho x_{t-1} + \ln(|\overline{u}| - e^{a^*}) + \frac{\beta}{1-\beta} \left(\frac{a^*}{1-\beta\rho} + \ln(|\overline{u}| - e^{a^*}) \right) + \frac{\beta\rho}{1-\beta\rho} a^* \right\}. \tag{17}$$

This problem has (after some rewriting) the associated first-order condition

$$\frac{e^{a^*}}{|\overline{u}| - e^{a^*}} = \frac{1}{1-\beta\rho} \Rightarrow$$

$$a^* = \ln\left(\frac{|\overline{u}|}{2-\beta\rho} \right).$$

This implies that the optimal wage payment is

$$w^* = -\ln(|\overline{u}| - e^{a^*}) = -\ln\left(|\overline{u}| \left(\frac{1-\beta\rho}{2-\beta\rho} \right) \right).$$

In the first-best equilibrium, then, the value of the firm is a linear function of the previous level of the dividend, while the wage and manager effort are constant. As either the rate of discounting β or the autoregressive parameter ρ

converge to zero the effort level converges to $\ln\left(\frac{|\overline{u}|}{2}\right)$, while the wage converges to $-\ln\left(\frac{|\overline{u}|}{2}\right)$. In these cases either the owner does not care about the future, or the manager does not contribute to future dividends through the past dividend level, and the manager's effort level is low. As β or ρ rise, however, the effort level rises and the wage rises to compensate.

7.2 Existence of the Second-Best Value Function

We seek to prove that a function v^* exists which solves:

$$v(x) = \max_{w,a^*} \left\{ \int_X (y - w(y)) g(y\,|x, a^*) dy + \beta \int_X v(y) g(y\,|x, a^*) dy \right\}$$

such that

$$a^* \in \arg\max_a \int_X u(w(y)) g(y \,|\, x, a) dy - h(a),$$

$$\int_X u(w(y)) g(y\,|x, a^*) dy - h(a^*) \geq \bar{u},$$

where X is the dividend space, y is the current (unobserved) dividend, x is the previous dividend, w is the wage paid to the manager, a^* is the manager's optimal action, and $\bar{u}$ is the manager's reservation utility. Note that for exposition our notation here differs slightly from the form in the text–in addition, we have replaced the current dividend density function dG with the density function gdy.

We assume that the first-order approach is valid; that is, the manager's preferences are such that the incentive compatibility condition (ICC)

$$a^* \in \arg\max_a \int_X u(w(y)) g(y \,|\, x, a) dy - h(a)$$

may be replaced with (cf. [18])

$$\int_X u(w(y)) \frac{\partial g(y \,|\, x, a^*)}{\partial a} dy - h'(a^*) = 0. \tag{ICC}$$

Assuming that the first-order approach is valid implies that an interior solution to the problem ICC exists, i.e. that the second-order condition satisfies

$$\int_X U(w(y)) \frac{\partial^2 g(y \,|\, x, a^*)}{\partial a^2} dy - h^{''}(a^*) < 0.$$

when the second-order condition exists. In order to ensure this, we must add the condition that the density function of current dividends $g\,(y \,|\, x, a)$ be at least twice-continuously differentiable in a.

From this we know immediately that the Implicit Function Theorem (IFT) applies around a^*; next we assume that $g\,(y \,|\, x, a)$ is at least twice-continuously differentiable in x so that we may write

$$\int_X U(w(y)) \frac{\partial g(y \mid x, a^*(x; w))}{\partial a} dy - h'(a^*(x; w)) = 0.$$

Note that since w takes as its argument the current dividend y, the optimal action a^* will not depend parametrically upon w, but rather functionally. By the IFT, however, we know that this functional dependence is one-to-one. Thus, the incentive compatibility condition defines the optimal action given the previous dividend and the wage function.

We may now write the dynamic programming problem of the owner as

$$v(x) = \max_w \left\{ \int_X (y - w(y)) g(y \,|x, a^*(x; w)) dy + \beta \int_X v(y) g(y \,|x, a^*(x; w)) dy \right\} \tag{18}$$

such that

$$\int_X u(w(y)) g(y \,|x, a^*(x; w)) dy - h(a^*(x; w)) \geq \overline{u} \tag{PC}$$

where PC is the participation constraint of the manager.

This problem has a straightforward solution. The current-period return function is bounded in w by assumption, and we also suppose that the conditional expected value of the current dividend is finite (i.e., $\int_X yg(y\,|x, a^*(x; w)dy < \infty$). The constraint PC is compact-valued, non-empty and continuous. As before (see equation 12) we define an operator $T : C_X^1 \to C_X^1$ by

$$T(v)(x) = \int_X (y - w^v(y)) g(y \,|x, a^*(x; w^v)) dy + \beta \int_X v(y) g(y \,|x, a^*(x; w^v)) dy \tag{19}$$

where w^v is the optimal solution to the problem (18) + (PC) for a given function v.

From the above considerations we know that the Theorem of the Maximum obtains–the T operator takes bounded continuous functions into bounded continuous functions. Furthermore, we can use Blackwell's sufficiency conditions to show that T is a contraction. Recall that if T is an operator taking bounded continuous functions into bounded continuous functions, then it is a contraction mapping if

1. T is monotonic, i.e. $T(v)(x) \leq T(w)(x)$ whenever $v(x) \leq w(x)\ \forall x$
2. T is 'sublinear', i.e. $T(v + c)(x) \leq T(v)(x) + \beta c\ \forall v, \beta \in (0, 1), c < 0$.

It is clear that the operator defined by (19) satisfies Blackwell's conditions, so that T is a contraction mapping. Hence, we know that a continuous function v^* exists which solves the owner's problem, and that

$$\lim_{n \to \infty} T^n(v_0) = v^*\ \forall v_0 \in C_X^1.$$

7.3 Second-Best Approximation Details

The second-best value function v^* was estimated using a single-layer feedforward neural network of 6 hidden units. Following the procedure outlined in Section 3, neural networks were fit to dividend grid data and then used to calculate the next iteration of the value function. Each neural network fit the grid data so that the sum-of-squared-error (SSE) between the estimated data and the actual data was less than 10^{-10}.

The dividend space X ranged from -2 to 20, and was divided into a grid $\{x_i\}$ of 30 equidistant points to be used as network input. These points were used to calculate the value function points $v^n(x_i)$. These points constituted the target vector for fitting the neural network. The actual grid data used to estimate the neural network varied from iteration to iteration. An adaptive grid was used to focus the neural network's attention on those points which were particularly hard to estimate, i.e. those grid points whose absolute errors were above the mean absolute error of the current iteration's estimate. Convergence of the value function was assumed when the largest absolute error between consecutive value function estimates was less than 10^{-3}. As mentioned in the text, the owner's discount factor β was set to 0.9, the autoregressive parameter ρ of the dividend process was also 0.9, and the reservation utility $\bar{u}$ was -2.

After the value function iteration had converged, grid data for the optimal wage function and the optimal effort function were generated. The optimal wage function was estimated using a single-layer feedforward neural network of 16 hidden units, and the SEE between the estimated and actual wage values was less than 10^{-4}. The optimal effort function required 6 hidden units, and the SSE between the estimated and actual effort values was less than 10^{-12}.

8 References

1. A. B. Abel. Stock prices under time-varying dividend risk. *Journal of Monetary Economics*, 22:375–393, 1988.
2. T. Bollerslev et al. Arch modeling in finance. *Journal of Econometrics*, 52:5–59, 1992.
3. W. A. Brock. Asset prices in a production economy. In J. J. McCall, editor, *The Economics of Information and Uncertainty*. University of Chicago Press and NBER, 1982.
4. T. E. Day and C. M. Lewis. Stock market volatility and the information content of stock index options. *Journal of Econometrics*, 52:267–287, 1992.
5. F. X. Diebold and M. Nerlove. The dynamics of exchange rate volatility: A multivariate latent factor arch model. *Journal of Applied Econometrics*, 4:1–21, 1989.
6. S. Durlauf. Locally interacting systems, coordination failure and the behavior of aggregate activity. Technical report, Department of Economics, Stanford University Working Paper, 1990.
7. R. F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.
8. E.F. Fama. The behavior of stock market prices. *Journal of Business*, 38:34–105, 1965.
9. A.R. Gallant et al. On fitting a recalcitrant series: The pound/dollar exchange rate 1974-83. Technical report, Department of Economics, Duke University Working Paper, 1989.
10. J. Geweke and S. Porter-Hudak. The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4):211–238, 1983.
11. A. Giovannini. Uncertainty and liquidity. *Journal of Monetary Economics*, 23:239–258, 1989.
12. C.W.J. Granger. The typical spectral shape of an economic variable. *Econometrica*, 34:150–161, 1966.
13. C.W.J. Granger. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14:227–238, 1980.
14. W. J. Den Haan and Scott Spear. Credit conditions, cross-sectional dispersion and arch effects in a dynamic equilibrium model. Technical report, Department of Economics, University of California at San Diego Working Paper, 1994.
15. M.L. Higgens and A. K. Bera. A class of nonlinear arch models. *International Economic Review*, 33(1):137–158, 1992.
16. K. Hornik et al. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Technical report, University of California-San Diego Working Paper, 1990.
17. J.R.M. Hosking. Fractional differencing. *Biometrika*, 68(1):165–176, 1981.
18. I. Jewitt. Justifying the first-order approach to principal-agent problems. *Econometrica*, 56(5):1177–1190, 1988.
19. K.L. Judd. *Numerical Methods in Economics*. MIT Press, 1998.
20. R.E. Lucas. Asset prices in an exchange economy. *Econometrica*, 46(6):1429–1445, 1978.

21. B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36:392–417, 1963.
22. B. Mandelbrot and J.W. van Ness. Fractional brownian motions, fractional noises and applications. *SIAM Review*, 10(4):422–437, 1968.
23. D.B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370, 1991.
24. P. Perron. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57(6):1361–1402, 1989.
25. W.P. Rogerson. The first-order approach to principal-agent problems. *Econometrica*, 53(6):1357–1367, 1985.
26. J.H. Stock. Measuring business cycle time. *Journal of Political Economy*, 95:1240–1261, 1987.
27. J.H. Stock. Estimating continuous time processes subject to time deformation. *Journal of the American Statistical Association*, 83:77–85, 1988.
28. N.L. Stokey and R.E. Lucas. *Recursive Methods in Economic Dynamics.* Cambridge: Harvard UP, 1989.
29. H. White. Learning in artificial neural networks: A statistical perspective. Technical report, University of California-San Diego, Working Paper, 1989.